GEOGRAPHIC AND INSTITUTIONAL DETERMINANTS OF REAL INCOME: A SPATIO-TEMPORAL SIMULTANEOUS EQUATION APPROACH*

by

Guyslain K. Ngeleza, Raymond J.G.M. Florax and
William A. Masters

Working Paper #06-15

December 2006

Department of Agricultural Economics

Purdue University

* The authors would like to thank the Economic Research Service of the United States Department of Agriculture (USDA/ERS) and the Purdue Research Foundation (PRF) for financial support. We also appreciate comments and suggestions of participants of the International Workshop on Spatial Econometrics and Statistics held in Rome, Italy, 25–27 May 2006.

GEOGRAPHIC AND INSTITUTIONAL DETERMINANTS OF REAL INCOME: A SPATIO-TEMPORAL SIMULTANEOUS EQUATION APPROACH

by

Guyslain K. Ngeleza[1], Raymond J.G.M. Florax[1,2] and William A. Masters[1]

[1] Dept. of Agricultural Economics, Purdue University
403 W. State Street, West Lafayette, IN 47907–2056, USA
Phone: +1 (765) 494–4300, Fax: +1 (765) 494–9176
E-mail: gngeleza@purdue.edu, rflorax@purdue.edu,
wmasters@purdue.edu

[2] Dept. of Spatial Economics, Vrije Universiteit
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Working Paper #06-15
December 2006

Abstract

This paper tests a series of prominent hypotheses regarding the determinants of per-capita income using a novel spatial econometric approach to control for spillovers among neighboring countries and for spatially correlated omitted variables. We use simultaneous equations to identify alternative channels through which country characteristics might affect income, and then test the robustness of those effects. We find support for both "institutionalist" and "geographic" determinants of income. A time-varying index of institutional quality has a strong independent effect on current income, but there is also a persistent effect of geographic factors such as seasonal frost, malaria transmission, and coastal location, which influence income through their links to agricultural output, health, urbanization and trade. The data cover 95 countries across the world from 1960 through 2002, which we use to construct a pooled dataset of nine 5-year averages centered on 1960, 1965, and so on through 2000. We use both limited and full information estimators, partly based on a generalized moments (GM) estimator for spatial autoregressive coefficients, allowing for spatial error correlation, correlation across equations, and the presence of spatially lagged dependent variables.

Key words: economic growth, geography, institutions, spatial econometrics, simultaneous equations
JEL codes: C31, C33, I18, O13, R12

GEOGRAPHIC AND INSTITUTIONAL DETERMINANTS OF ECONOMIC GROWTH: A SPATIO-TEMPORAL SIMULTANEOUS EQUATION APPROACH

1. Introduction

Real income per person varies widely around the world. Most of that variation is associated with differences across countries, rather than within them (Sala-i-Martin 2006). What accounts for this "country effect" on productivity and well-being? In this paper we test some of the most influential theories using a new econometric estimator, allowing for a wide range of neighborhood effects in space and time.

Following Douglass North (1990) and others, economists have long argued that income depends primarily on economic and political institutions. This view emphasizes the role of property rights, market infrastructure and price incentives as the key cause of differences in investment and economic growth. These institutions may correspond to national government policies, but they may arise and spread in other ways as well. An alternative approach championed in recent years by Jared Diamond (1997) and particularly by Jeffrey Sachs (2001) uses location-specific geographic and technological factors to explain income differences. This approach argues that geographic obstacles to improving public health, agricultural productivity and transportation infrastructure could explain cross-country differences in average incomes, and perhaps also help account for cross-country differences in economic institutions.

The policy implications of the two views are starkly different. The "institutions first" approach implies that countries can all converge to a common level of income, provided that they adopt similarly favorable institutions. The "technology first" view argues that those institutions will not have similar payoffs everywhere, leaving the poorest regions in need of exogenous injections of public investment for R&D and infrastructure to overcome their geographic handicaps. Intervention might be most important in sectors where technologies are location-specific, particularly in agriculture, where different locations require different crop and livestock techniques, and in public health, where different locations face different kinds of disease pressure.

A principal challenge in testing among competing hypotheses is endogeneity: any observed correlation between institutions and income could be due to reverse causality, or to omitted variables that affect both of them. Using a system of equations approach to account for reverse causality some of the most widely cited results are due to Rodrik, Subramanian and Trebbi (2004), Easterly and Levine (2003) and Acemoglu, Johnson and Robinson (2001). These studies find that geographic location is correlated with income only through its influence on institutions, and geographic location as such has no additional effect. But researchers using other variables and specifications have found different results. For example, Gundlach (2004) finds a large and robust influence of the location-specific degree of malaria transmission, independently of a country's institutions.

A second fundamental challenge in testing across countries is spatial correlation: there are obvious geographical clusters of rich countries and poor ones. Geographic clustering could be due to spatially correlated attributes such as climate or access to transport, or to interactions among neighbors such as trade or migration; for recent reviews, see Magrini (2004) and Abreu, de Groot and Florax (2005a). We have data for some geographic attributes and for some interactions among countries, but there are inevitably omitted variables of both types which could account for such geographic correlations as the synchronized growth fluctuations in Latin American and elsewhere documented by Temple (1999). In this paper we build on the spatial estimator developed by Kelejian and Prucha (2004) to control for very general kinds of neighborhood and spatial spillover effects, while allowing for endogeneity of key regressors. Doing so raises the bar for each hypothesis, by testing them against a wider range of alternative processes.

The remainder of this paper is organized as follows. Section 2 introduces a particular simultaneous equation model accounting for several possible channels by which institutions and technology could influence income. Section 3 presents the data, and provides an exploratory empirical assessment of the dynamics over space and time of the key variables in the system. In Section 4, we compare the estimation results of a non-spatial version of the system to the results based on the Kelejian and Prucha (2004) estimator allowing for spatial error autocorrelation, correlation across equations, and the presence of spatially lagged dependent variables. Section 5 provides conclusions.

2. Determinants of income across countries: a spatial system of equations approach

The modern literature on the empirics of economic growth across countries begins with Baumol (1986), who established the tendency of countries to converge towards a common income level. Such convergence was predicted by growth models with diminishing returns to capital, and was found by a very large number of studies using the specification pioneered by Barro (1991). Typically, convergence was found to be conditional on a number of factors, including both geographic and technological variables as well as institutional or policy measures (see e.g., Sala-i-Martin 1997), but the estimated coefficients varied widely and their interpretation remained controversial (see Abreu, de Groot and Florax 2005b, and Dobson, Ramlogan and Strobl 2006, for an overview). Following Hall and Jones (1999), attention shifted to the determinants of income levels, focusing particularly on the development of new identification strategies such as those introduced by Acemoglu, Johnson and Robinson (2001) to account for the endogeneity of economic institutions and policy choices.

In this paper, we want to allow for different types of spatial autocorrelation processes to affect a variety of endogenous variables, and also to affect income through other means. For this purpose, we adopt an explicit three-stage least squares approach, with panel data in a system of simultaneous equations. By identifying the entire system, the role of each possibly endogenous determinant of income is tested through an association with particular exogenous variables. Our identification strategy rests on that exogeneity, together with the exclusion restrictions by which those variables are tied to particular development channels (Klein and Vella 2005). These identifying assumptions are plausible but are not tested here. Our goal is to posit a relatively large and quite general representative system, and leave more general testing of its particular specification to future work.

The particular system of equations we use specifies six endogenous variables that jointly influence income. The endogenous variables are: *agricultural output,* as measured by the Food and Agricultural Organization's (FAO) index of net production at international prices; *infant mortality*, as estimated by the World Health Organization, which we use as a measure of general health; *schooling,* from the Barro-Lee measure of average educational attainment for the population; *institutional quality,* using a combination of measures from Freedom House and the

International Country Risk Guide (ICRG); *urbanization,* the fraction of the population in towns and cities reported to the FAO; *trade*, using the sum of exports plus imports as a fraction of GDP, from the Penn World Tables (PWT) 6.2; and finally *income*, using real GDP per capita, chain indexed, also from version 6.2 of the PWT.

Each of our endogenous variables has been widely used in the growth literature. Together, they capture a broad range of potential growth mechanisms, whose significance is econometrically identified and tested across seven equations using ten variables, which we assume to be exogenous. The exogenous variables can be grouped in five different categories: farmland and climate, disease ecology, social history, coastal location, country size.

The set of farmland and climate variables helps identify the potential influence of agricultural output. The specific variables we use are the FAO estimate of *agricultural land* area, an estimate of average *land quality* from the United States Department of Agriculture (USDA), and data on prevalence of *frost in winter* and *annual rainfall* from the Intergovernmental Panel on Climate Change (IPCC). These variables are included as potential determinants of agricultural production, but are plausibly excluded from having any significant direct effect on economic activity outside of agriculture.

The disease ecology variable helps identify the potential influence of disease transmission on health, through either labor productivity in agriculture or through infant mortality to non-agricultural activities. The specific variable we use is *malaria ecology* from Kiszewski et al. (2004), which captures the ease with which a mosquito-borne disease would spread from person to person, whether or not the disease is actually present. This variable is expected to have an influence on agricultural productivity and on infant mortality, but it is excluded from having any direct effect elsewhere.

A set of social history variables helps identify the role of institutional quality. The specific measures we use are the percentages of the population that are *Protestant*, *Catholic* or *Muslim*, to capture the degree to which a country has been influenced by world cultures that spread through migration and military conquest out of Northern Europe, Southern Europe, or the Middle East respectively. These variables are included as potential determinants of institutional quality, and excluded elsewhere.

The coastal location variable helps identify the potential impact on growth of either agglomeration in cities or international integration through trade. The specific *coastal* variable used here is percentage of the population located within 100 km of the ocean or a navigable river. This variable is included only as a determinant of agglomeration and of international trade.

Finally, a variable reflecting the size of the country is used only as a conditioning variable with respect to trade. Our specific variable is *total population,* included here because a smaller country will have a larger fraction of its transactions classified as "international", simply because of where its borders are drawn. This variable is excluded from other equations, on the grounds that researchers have found very limited scale effects in most income regressions.

The resulting system of equations is recapitulated below. The implied exogeneity and exclusion restrictions are plausible but, as noted above, specification and robustness tests are left to future work. Here, our goal is to estimate this representative system taking into account neighborhood effects through spatially correlated omitted variables and spatial spillover effects from the dependent variables. The system that we use is chosen primarily for its size and generality, capturing a wide range of potential growth mechanisms and linkages. Note that we focus here on the cross-sectional properties of the panel. Time dummies are used for each five-year period, to absorb any global trends in each equation; future work might focus on temporal dynamics.

The first equation of the system uses ecological variables to identify exogenous determinants of agricultural output:

$$\begin{aligned} agoutput_{it} = \alpha_1 + \beta_{11} agland_{it} + \beta_{12} landqual_i + \beta_{13} frost_i + \beta_{14} rainfall_i \\ + \beta_{15} malaria_i + \delta_{1t} + \varepsilon_{1it}. \end{aligned} \tag{1}$$

In this equation agricultural production per capita is a function of land area per capita, the soil quality of that land, the prevalence of seasonal frost and total rainfall, plus the ecological index of malaria transmissibility. These factors could be associated with exogenously higher agricultural output, which in turn could influence economy-wide income through a number of mechanisms, both positive (e.g., Mellor and Johnston 1961) and negative (e.g., Matsuyama 1992).

The next two equations use malaria transmission to identify exogenous determinants of human capital. This is done first for health, as measured by infant mortality:

$$imrate_{it} = \alpha_2 + \beta_{21} income_{it} + \beta_{22} malaria_i + \delta_{2t} + \varepsilon_{2it} \tag{2}$$

In equation (2), infant mortality is subject to feedback effects from income, and potentially also to an exogenous effect from malarial ecology. An exogenously driven change in health could matter for income in many ways, including acceleration of investment in schooling as captured in the next equation:

$$schooling_{it} = \alpha_3 + \beta_{31} income_{it} + \beta_{12} imrate_{it} + \delta_{3t} + \varepsilon_{3it} \tag{3}$$

In equation (3) there are no exogenous variables. We include this equation only to identify a possible channel for health to influence growth through education as opposed to other mechanisms.

The following equation uses social history to identify exogenous determinants of a country's institutions:

$$\begin{aligned} institqual_{it} = \alpha_4 + \beta_{41} income_{it} + \beta_{42} imrate_{it} + \beta_{43} pctcath_i + \beta_{14} pctprot_i \\ + \beta_{45} pctmus_i + \delta_{4t} + \varepsilon_{4it} \end{aligned} \tag{4}$$

Equation (4) links our Freedom House-ICRG index of institutional quality to economy-wide income, the infant mortality rate as a measure of human capital, and social history defined using prevalence of three global religions that were spread from Europe and the Middle East across Asia, Africa and Latin America through migration and military conquest.

The next equation uses coastal location as an exogenous driver of opportunities for specialization and exchange in towns and cities:

$$urbanization_{it} = \alpha_5 + \beta_{51} income_{it} + \beta_{52} agoutput_{it} + \beta_{53} coastal_i + \delta_{5t} + \varepsilon_{5it} \tag{5}$$

Equation (5) allows urbanization to be driven by feedback from income and also from agricultural output, as well as access to coasts or navigable rivers.

An alternative route to specialization is captured in the next equation, which identifies other determinants of international trade:

$$trade_{it} = \alpha_6 + \beta_{61} income_{it} + \beta_{62} coastal_i + \beta_{63} population_{it} + \delta_{6t} + \varepsilon_{6it} \tag{6}$$

In equation (6), trade can be driven by economy-wide income, coastal location and population size.

The last equation brings the endogenous variables together, with no additional exogenous variables.

$$\begin{aligned} income_{it} = \alpha_7 &+ \beta_{71} agoutput_{it} + \beta_{72} imrate_{it} + \beta_{73} schooling_{it} \\ &+ \beta_{74} instqual_{it} + \beta_{75} urbanization_{it} + \beta_{76} trade_{it} + \delta_{7t} + \varepsilon_{7it} \end{aligned} \tag{7}$$

This system of equations can be estimated using 3SLS, but the results are likely to be biased and/or inefficient due to spatial processes beyond those captured in the regressors. The equations may share spatially autocorrelated errors due to spatially correlated omitted variables, to spatially correlated measurement error, or to interaction among neighboring countries as detailed by Anselin (2003). In this paper we account for spatially correlated residuals in a system of equations by allowing each endogenous variable to be subject to both spatial dependence and also to a spatial autoregressive process in the error term (i.e., a spatial ARAR model). For this we utilize a recently developed full information estimator based on Instrumental Variable (IV) and General Moments (GM) estimators, which simultaneously allows for correlation across equations (Kelejian and Prucha 2004).[1] Here, we start with the naïve three-stage least squares approach, and then compare these results to the estimates allowing for the potential influence of spatial spillovers and spatially-correlated omitted variables.

3. Data and some exploratory results

For all time-variant data, we use observations at five-year intervals, around 1960, 1965, and so on through 2000. In most cases these are an average of five annual observations centered on the year indicated (that is, 1963–67 for 1965, 1968–72 for 1970, and so forth), although only three years are available to represent 1960 (that is, 1960–62) and only three years to represent 2000 (that is, 1997–2000). For the Barro-Lee (2001) data on schooling and also the UN data on infant mortality, single-year observations are used at the corresponding five-year intervals.

[1] Kelejian and Prucha (2005) developed an extended estimator that incorporates heteroskedasticity as well, which can be incorporated in future work.

The agriculture data are updated from the dataset in Masters and Wiebe (2000), using FAO (2004) data for each year's level of agricultural output (expressed in real international dollars of the year 2000) and land used in agriculture (expressed in thousands of hectares). The land quality index is from USDA (2005), reporting the fraction of a country's agricultural land that is reported to be in the top three categories of suitability for agriculture in the World Soil Resources classification scheme, as reported in NRCS (1999). Agricultural land is defined broadly here, to include 'cropland' and 'cropland plus natural mosaic' from the International Geosphere-Biosphere Programme classification (USGS 1999).

Climatic data were compiled by Masters and McMillan (2001) from data published by the International Panel on Climate Change (IPCC 1999). Frost prevalence refers to the proportion of a country's land receiving five or more frost days in that country's winter, defined as December through February in the Northern hemisphere and June through August in the Southern hemisphere. The raw data for this computation were the IPCC's estimated average number of frost-days per month over the 1961–90 period, across 0.5-degree cells for all land mass except Antarctica, interpolated from station observations. Rainfall is average total annual precipitation for each grid cell averaged over the country's landmass. The country aggregation is based on the CRU TS 2.0 gridded dataset (Mitchell et al. 2003).

Economic data are drawn from the Penn World Tables 6.2 for national income (real GDP per capita, chain indexed, in 2000 US dollars) and for the trade share (exports plus imports as a fraction of GDP). Urbanization is drawn from the World Development Indicators online, as the percentage of the population in urban areas. Data on schooling are drawn from Barro and Lee (2001), from which we use the average number of years of total schooling in the population over age 15. Data on infant mortality rates are drawn from United Nations Population Statistics, and our data on long-run social history are the percentages of the population estimated to be Protestant, Catholic, and Muslim from the Barro-Lee dataset.

Our malaria ecology variable is from Kiszewski et al. (2004), and represents an index of the ease with which a given infection would be transmitted, independently of whether the infection is present. The index is constructed from the physiological characteristics of each region's dominant mosquito species, combined with temperature data that determine how long a malaria parasite could survive during transmission from person to person. These factors are largely independent of a country's economic activity or its anti-malarial efforts. Most importantly, the index does not include data on the density of mosquitoes or the prevalence of infection, both of which can be reduced in an otherwise malarial region.

Our variable for the quality of national institutions is a time-varying index, constructed by us from data reported by Freedom House (2005) and International Country Risk Services (ICRG 2006). The Freedom House data is an average of their measures for a country's political rights and civil liberties, whereas the ICRG index is an average of their measures for a country's degree of corruption, military in politics, religion in politics, law and order, and democratic accountability. Data from the two sources are rescaled for comparability, and combined to construct a continuous time series from 1960 to 2000.

Overall, the dataset comprises nine five-year averages pertaining to 1960 through 2000 for 95 countries of the world. Variable definitions and descriptive statistics are provided in Table 1, with a complete list of countries provided in the appendix. Our coverage includes all of North and South America except for Belize, Suriname, French Guiana and some islands in the Caribbean. In order to build consistent data series several African countries, such as Morocco, Libya, Chad, Ethiopia, Nigeria and Chad, could not be included in the sample. Switzerland and Germany as well as most of the Central and Eastern European countries are excluded, as well as Russia, Mongolia and some smaller countries in South-East Asia.

The geographic distance between countries is captured through a spatial weights matrix, which is defined a priori and exogenously on the basis of arc-distances between the geographical midpoints of the countries considered. It is an inverse-distance matrix where elements are coded $1/d_{ij}$ if the distance between countries $d_{ij} \leq 2{,}500$ miles. Following convention, we standardize by enforcing row sums to be equal to one and the diagonal elements set to zero (see e.g., Bell and Bockstael 2000, for an explanation). The resulting spatial weight matrix for a single time slice has dimension 95, with 17% of the weights being nonzero. The minimum and maximum number of links between countries is 1 and 26, respectively, with an average of 16. The minimum cutoff distance required to ensure that each country would be linked to at least one other country would have been 1,812 miles. In our weight matrix, the connectivity structure is such that there is no direct link between America and Europe, although some countries in South America are directly linked to Africa. The weight matrix for the pooled data set is defined as a 855×855 block diagonal matrix, with the sequence of nine 95×95 matrices on the diagonal. This implies that we assume spatial autocorrelation to be strictly contemporaneous.

Figure 1 represents key information for 2000 in choropleth maps, specifically for the dependent variables in the system of equations developed in Section 2. GDP per capita, measured in constant US dollars of 2000, is highest in North America, Europe, Australia and Japan, and is relatively low in South America, Asia and especially the African continent. The spatial distribution of per capita agricultural output is very similar, although some of the European countries have a less pronounced position, and the agricultural output levels are dramatically low in Africa. The spatial distribution of the trade share in GDP shows a much more scattered picture. Apart from a city-state such as Singapore, which is hard to see on the map, countries with relatively high trade shares include Guyana and Malaysia as well as Ireland, Belgium and the Netherlands. Infant mortality rates are highest in India, Pakistan, Iraq and Sub-Saharan Africa, and comparatively low in the industrialized economies of North America, Europe and Australasia. The spatial distribution of institutional quality exhibits a concentration of high quality institutions in North America, Northern Europe and Australasia, and Southern Europe constitutes an intermediate zone. For schooling a spatial pattern similar to institutional quality arises, although in most countries in South America schooling duration is above average as well. The spatial distribution of the level of urbanization, defined as the percentage of the population living in urban areas, is much more uniform. Except for Southeast Asia and Sub-Saharan Africa, the level of urbanization is generally greater than 50% throughout the world.

Figure 2 summarizes the level and changes in spatial clustering for the endogenous variables using Moran's *I* statistic, defined as the degree of correlation between each country's value and that of its neighbors.[2] Global GDP per capita values have a high degree of spatial clustering at the start of the period, suggesting strong neighborhood effects, with a small further

[2] With a standardized weights matrix Moran's *I* is defined as:

$$I = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x}) / \sum_{i=1}^{n}(x_i - \bar{x})^2$$

where the variable x is measured in deviations from its mean, and w_{ij} are the elements of the weights matrix. The expected value of Moran's I equals $-1/(n-1)$ under the null hypothesis of no spatial autocorrelation, which is approximately –0.01 for our sample and signals a random spatial allocation of the attribute values contained in x. We use the normal distribution assumption for statistical inference. Extensive details and principles for statistical inference are available in Cliff and Ord (1981) and Tiefelsdorf (2000).

increase in clustering during the 1970s and 1980s. Agricultural output, in contrast, exhibits a much lower degree of spatial autocorrelation, and a greater increase especially in the 1980s. The trade share has even lower spatial clustering, which increases over time but remains very close to zero. Infant mortality rates start highly clustered and become even more so, exhibiting the highest degree of spatial autocorrelation during the entire period. Schooling and institutional quality start with similar levels of clustering in the early 1960s, but schooling becomes much more clustered while institutional quality does not. In fact, spatial clustering of institutional quality actually declines slightly from its peak in the 1980s through the 1990s. Finally, urbanization has a relatively high degree of spatial clustering, and shows a small increase except for 1995 and 2000.

Figure 3 provides some more detail with respect to the spatial distribution of the seven dependent variables in Moran scatterplots, for the latest available period. These charts show the standardized value of each country's variable x_i against its spatial lag, which equals the spatially weighted average of the x_j-values with the set of neighbors being defined through the i-th row of the weights matrix. It aids in identifying local clusters of spatial correlation, spatial non-stationarity and outliers, and the gradient of the trend line equals the Moran's I coefficient (see Anselin 1996 for details).

In Figure 3, the Moran scatterplot for GDP per capita shows a strong clustering of countries in the lower-left quadrant, which are low-income countries surrounded by countries with similarly low per capita incomes. A few low-income countries, however, are in the upper-left quadrant, meaning that their neighbors actually have above-average incomes. In contrast, many of the high-income countries are in neighborhoods with above-average income. The outliers (as judged by the 2σ-rule) are the US, which is surrounded by neighbors with average per capita income, and Norway, which is surrounded by above-average per capita income neighbors. The scatterplot for agricultural output is similar to the plot for GDP per capita although the extent of spatial clustering is smaller. Outliers are New Zealand, Australia, Denmark, and Ireland. As mentioned above there is no significant spatial clustering in trade, and Singapore is the extreme outlier. The scatterplot for infant mortality clearly shows two separate clusters, of which one comprise most of the countries located in Africa. Institutional quality, schooling and urbanization show a similar degree of spatial clustering, without any obvious outliers.

4. Econometric method and estimation results

In two recent papers reviewing the economic growth literature Abreu, de Groot and Florax (2005a,b) stress two important implications of earlier work. First, in a quantitative analysis of over 600 estimates drawn from nearly 50 convergence studies they find that correcting for endogeneity in the explanatory variables results in significantly higher estimates of the rate of convergence. This is in line with earlier findings of Cho (1996) and Caselli, Esquivel and Lefort (1996). In addition, they document that the use of panel data and concurrent corrections for unobserved heterogeneity in technology levels and/or steady states leads to substantially higher rates of convergence, which is reinforced by the results of Dobson, Ramlogan and Strobl (2006). Second, the Abreu et al. reviews of the spatial econometric literature dealing with (regional) economic growth shows that this literature has not yet established a strong link to prevalent economic growth theories, and it has a tendency to restrict the modeling of spatial spillover processes to either a spatial lag or a spatial error model, eventually in combination with spatial regimes to account for non-stationarity in the mean and variance. Only recently have spatial methods been more rigorously applied, as in Ertur and Koch (2005) and Fingleton and López-Bazo (2006).

In the current paper we follow the approach outlined in Kelejian and Prucha (2004) and use a spatial econometric specification that is less restrictive than previous work in terms of spatial correlation, and accommodates endogeneity at the same time. In terms of spatial autocorrelation, the specification allows for spatial spillover effects through the dependent variable as well as for a spatial autoregressive error structure. This specification is known as the spatial ARAR model. For a single equation this specification reads as:

$$\begin{aligned} y &= \rho W y + X\beta + \varepsilon, \\ \varepsilon &= \lambda W \varepsilon + \mu, \end{aligned} \tag{8}$$

where y is an $(n\times1)$ vector of observations on the dependent variable, X an $(n\times k)$ matrix of non-stochastic regressors, W an $(n\times n)$ spatial weights matrix that represents the topology of the spatial system, μ an $(n\times1)$ vector of iid errors, β a $(k\times1)$ vector of regression coefficients, and ρ

and λ are spatial autoregressive parameters. Substitution and rearrangement of terms in equation (7) leads to:

$$y = (I - \rho W)^{-1}\left(X\beta + (I - \lambda W)^{-1}\mu\right), \tag{9}$$

which shows that equation (7) implies a rather complex form of spatial autocorrelation evoked by nested spatial multiplier processes pertaining to the observable and the non-observable part of the model (see also Anselin 2003). The spatial complexity of the model notwithstanding, testing for spatial autocorrelation is rather straightforward and can be based on a Lagrange Multiplier test for which the asymptotic distribution has been derived in a maximum likelihood framework. This test is generally known as the SARMA test but since Lagrange Multiplier tests cannot distinguish between locally equivalent autoregressive (AR) and moving average (MA) processes (Godfrey 1988) the SARMA test can also be used to detect an ARAR process.[3]

Instead of a purely cross-sectional dataset, we use a panel dataset comprising nine time slices centered on 1960, 1965, etc. through 2000. We do not investigate the temporal dynamics and associated serial autocorrelation, but simply treat the data as independent replications of the cross-sectional data. We do, however, include fixed effects for the different time periods, thus accommodating a possible time trend. Given that some data offer yearly observations, richer models incorporating spatio-temporal dynamics are feasible, but we leave those for future research (see Anselin, Le Gallo and Jayet 2006).

A distinct advantage of the Kelejian and Prucha (2004) systems approach is that it explicitly allows for endogeneity to be taken into account. The endogeneity is not necessarily restricted to spatial spillover effects, but it can also include the usual system feedback effects. Kelejian and Prucha (2004) derive a full information generalized spatial systems estimator (GS3SLS) in a sequential estimation procedure using limited information IV and GM estimation to provide initial estimates of the spatial autoregressive parameters. The set-up and the estimators involved are described concisely as follows.

[3] Anselin and Kelejian (1997) discuss testing for spatial autocorrelation in a model with endogenous regressors, where the endogeneity is caused by systems feedbacks or by spatial interaction of an endogenous variable. In the empirical application we initially use OLS based tests although this ignores the endogeneity of some of the regressors. Testing for spatial autocorrelation can also be based on the general results for Moran's *I* in Kelejian and Prucha (2001).

Consider a simultaneous system of m spatially interrelated cross-sectional equations indexed by j (= 1, 2, …, m) and defined as:

$$Y = \overline{Y}\mathrm{P} + Y\Gamma + X\mathrm{B} + U \,, \tag{10}$$

where $Y = (y_1, y_2, \ldots, y_m)$ with y_j as the (n×1) vector of observations on the dependent variable, $\overline{Y} = (\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_m)$ has the same dimension and contains the spatial lags of the endogenous variables defined as $\overline{y}_j = Wy_j$, $X = (x_1, x_2, \ldots, x_k)$ with x_l as the (n×1) vector of observations on the exogenous variable l, and $U = (u_1, u_2, \ldots, u_m)$ where u_j is the vector of errors in the jth equation. Further, W is an (n×n) spatial weights matrix of known constants, and P is an (m×m), Γ an (m×m) and B a (k×m) parameter matrix. In addition to the spatial spillovers in the endogenous variables the errors are also allowed to include a spatial autoregressive process:

$$U = \overline{U}\Lambda + E \,, \tag{11}$$

with $E = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)$ where ε_j denotes the (n×1) vector of innovations. Analogous to the spatial lag operations above, $\overline{U} = (\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_m)$ are the spatially correlated errors with $\overline{u}_j = W\overline{u}_j$, and the spatial autoregressive parameters are given by $\Lambda = \mathrm{diag}_{j=1}^{m}(\lambda_j)$.

One should note that the coefficient matrix P referring to the spatially lagged endogenous variables is not necessarily diagonal, and hence the specification allows for the jth endogenous variable to depend on its own spatial lag as well as on spatial lags of other endogenous variables. We leave this generalization to future work. The coefficient matrix Λ is also assumed to be diagonal, implying that the errors are spatially correlated within an equation, but they are not spatially correlated across equations.[4] The generality of the systems approach and the suggested estimator is also evident from the fact that the exogenous regressors are allowed to depend on n, and hence form triangular arrays, which implies that the specification may also contain spatially lagged exogenous variables (Kelejian and Prucha 2004, p. 30). As a final observation we note that using the feasible GS3SLS estimator makes Wald tests available to test restrictions on the (spatial autoregressive) parameters.[5]

[4] The GS3SLS estimator allows for error correlation across equations, but this correlation does not have a spatial dimension.

[5] As far as the spatial variables are concerned, this is only feasible for the spatially lagged endogenous variables and eventually the spatially lagged exogenous variables. A Wald test on spatially autocorrelated errors is not possible, because the values of λ_j are merely used in the Cochrane-Orcutt transformation. The latter can be tested using

In order to determine the marginal effects of changes in the exogenous variables we use the notation and the line of reasoning introduced in Kelejian and Prucha (2004, pp. 30-31). Define $\mathrm{y} = \mathrm{vec}(Y)$ and corresponding operations to define $\bar{\mathrm{y}}, \mathrm{x}, \mathrm{u}, \bar{\mathrm{u}}$ and ε. Given that $\bar{\mathrm{y}} = (I_m \otimes W)\mathrm{y}$, the system defined by (9) and (10) can be written as:

$$\begin{aligned} \mathrm{y} &= \Gamma^* \mathrm{y} + \mathrm{B}^* \mathrm{x} + \mathrm{u}, \\ \mathrm{u} &= \Lambda^* \mathrm{u} + \varepsilon, \end{aligned} \tag{12}$$

where $\Gamma^* = (\Gamma' \otimes I_n) + (\mathrm{P}' \otimes W)$, $\mathrm{B}^* = \mathrm{B}' \otimes I_n$ and $\Lambda^* = \Lambda \otimes W = \mathrm{diag}_{j=1}^{m}(\lambda_j W)$. The reduced form of (11) then follows from rearranging terms as:

$$\mathrm{y} = (I_{nm} - \Gamma^*)^{-1}[\mathrm{B}^* \mathrm{x} + (I_{nm} - \Lambda^*)^{-1} \varepsilon], \tag{13}$$

where I_{nm} has dimension ($nm \times nm$). Marginal effects of changes in more or more of the exogenous variables follow from:

$$\frac{\partial \mathrm{y}}{\partial \mathrm{x}'} = (I_{nm} - \Gamma^*)^{-1} \mathrm{B}^* = [I_{nm} - (\Gamma' \otimes I_n) - (\mathrm{P}' \otimes W)]^{-1} \mathrm{B}^*. \tag{14}$$

This equation shows that the impact of a shock to one or more of the exogenous factors leads to spatial feedback via the endogenous regressors (through the term $\Gamma' \otimes I_n$), and depends on the geographical location and the spatial connectedness of the place where the exogenous shock occurs (which is contained in the term $\mathrm{P}' \otimes W$). The weights matrix W defines the extent of each country's neighborhood, and hence the limits of these spatial feedback effects. In our application, the definition of neighborhood is extremely broad to capturing a very wide range of spillovers, as all countries within a 2500 miles radius are linked to each other. Further work could test more restrictive specifications.[6]

Moran's I (see Kelejian and Prucha 2001), or the Lagrange Multiplier principle (see Anselin and Kelejian 1997). See also footnote 2.

[6] An alternative approach uses direct representation of a distance decay process for spatial spillovers, in a parametric or non-parametric fashion (see, e.g., Conley and Ligon 2002). Some work has also pursued endogenizing the spatial weights matrix (Kelejian and Prucha 2005). However, neither approach can circumvent the occurrence and relevance of the Modifiable Areal Unit Problem (MAUP; see, e.g., Anselin 1988).

In a concise form we can write (9) and (10) as a system of cross-sectional equations indexed by j = (1, 2, …, m):

$$y_j = Z_j\delta_j + u_j,$$
$$u_j = \lambda_j W u_j + \varepsilon_j, \tag{15}$$

where $Z_j = (\bar{Y}_j, Y_j, X_j)$ and $\delta_j = (\rho_j', \gamma_j', \beta_j')'$. The full information estimator derived in Kelejian and Prucha (2004) is obtained in the following four steps:

1. Apply 2SLS to each equation and estimate δ_j as $\tilde{\delta}_j = (\tilde{Z}_j' Z_j)^{-1}\tilde{Z}_j' y_j$, where $\tilde{Z}_j = P_H Z_j$, $P_H = H(H'H)^{-1}H'$ and H is a matrix of instruments formed as a subset of the linearly independent columns of $(X, WX, W^2X, ...)$.
2. Based on $\tilde{\delta}_j$, compute the 2SLS residuals $\tilde{u}_j = y_j - Z_j\tilde{\delta}_j$ and use the generalized moments procedure suggested in Kelejian and Prucha (1999) to estimate λ_j, the spatial autoregressive parameter of the error process for each equation.
3. Use a Cochrane-Orcutt transformation to define the suitably transformed variables $Z_j^* = Z_j - \tilde{\rho}_j W Z_j$ and $y_j^* = y_j - \tilde{\rho}_j W y_j$, and apply a feasible generalized spatial 2SLS estimator (FGS2SLS) to obtain $\hat{\delta}_j^{F2SLS} = (\hat{Z}_j'^{*} Z_j^*)^{-1}\hat{Z}_j'^{*} y_j^*$ where $\hat{Z}_j^* = P_H Z_j^*$.
4. Stack the equations as $y^* = Z^*\delta + \varepsilon$, where $y^* = (y_1'^{*}, y_2'^{*}, \ldots, y_m'^{*})'$, $Z^* = \mathrm{diag}_{j=1}^{m}(Z_j^*)$, and $\delta = (\delta_1', \delta_2', \ldots, \delta_m')'$. Obtain the full information results by using the feasible GS3SLS estimator to calculate $\hat{\delta}^{F3SLS} = (\hat{Z}'^{*}(\hat{\Sigma}^{-1}\otimes I_n)Z^*)^{-1}\hat{Z}'^{*}(\hat{\Sigma}^{-1}\otimes I_n)y^*$, where $\hat{\Sigma}$ is estimated as an (m×m) matrix whose j,l-th element is $\hat{\sigma}_{jl} = n^{-1}\tilde{\varepsilon}_j'\tilde{\varepsilon}_l$ with $\tilde{\varepsilon}_j = y_j^* - Z_j^*\hat{\delta}_j^{F2SLS}$. Kelejian and Prucha (2004) prove that the small sample distribution of the FGS3SLS estimator can be approximated by $\hat{\delta}^{F3SLS} \stackrel{.}{\sim} \mathrm{N}(\delta, [\hat{Z}'^{*}(\hat{\Sigma}^{-1}\otimes I_n)\hat{Z}^*]^{-1})$.

The asymptotic properties of the above estimator critically depend on the assumption of homoskedastic innovations. In future work we will extend the application to the ARAR estimator allowing for heteroskedasticity along the lines developed in Kelejian and Prucha (2005).

We now turn to the estimation results for the system of equations developed in Section 2. The results are generated using the same spatial weights matrix throughout the entire model, and we account only for spatially lagged dependent rather than the more general spatially lagged endogenous variables (across equations). We do also not incorporate spatially lagged exogenous variables. We follow Kelejian and Robinson (1993) and define the instruments as the linearly independent exogenous variables and their first-order spatial lags, although alternatives would be available as well (see Lee 2003). Table 2 provides the results for an equation-by-equation estimation using OLS and includes several (spatial) diagnostic test results. Table 3 presents the 3SLS results, which account for endogeneity but not spatial effects, while Table 4 presents the full information results using the feasible GS3SLS estimator discussed above.

Table 2 is provided for comparative purposes. This naïve specification, without any control for endogeneity or spatial lags, shows how income is closely correlated with a number of endogenous regressors, notably infant mortality, schooling, institutional quality and urbanization. Each of them is in turn also correlated with income, when controlling for various other significant determinants. The misspecification test results shown here are also only heuristic, since they are derived without accounting for the endogeneity of some of the regressors. The condition number shows that multicollinearity does not impair the results. The results for the Jarque-Bera test indicate that the null hypothesis of normally distributed errors is rejected for nearly all equations. This provides another reason for interpreting the Lagrange Multiplier diagnostics cautiously. It does not, however, have any major implications for the systems estimator, because the estimator does not require the disturbances to be normal. The Breusch-Pagan test results, with random coefficient variation as the alternative hypothesis, show that homoskedasticity is rejected in all. This implies that it is highly relevant to address this issue in future work. The spatial diagnostics are fairly mixed. For six out of seven equations there is evidence that a higher-order model is appropriate (in particular for the equations pertaining to agriculture output, infant mortality, schooling, urbanization, trade and income). There is, however, no clear indication of spatial autocorrelation for the institutional quality equation.

Tables 3 and 4 contain the estimation results for the systems estimators. Table 3 accounts for endogeneity using 3SLS, whereas Table 4 accounts for both endogeneity and neighborhood effects using the spatial ARAR model. Briefly, the results of Table 3 can be summarized as follows. Unlike the naïve OLS regression, in a system context per-capita income is not correlated with institutional quality. Income has a strong and significant links from infant mortality, and agricultural output. Per capita agricultural output, in turn, is mainly determined by the availability of agricultural land and land quality; the prevalence of winter frost has the same significant effect as in Masters and McMillan (2001), while disease ecology (malaria) also has the expected effect. Infant mortality is mainly linked to income; controlling for income and the time dummies disease ecology show also strong correlation with infant mortality. Schooling and institutional quality are both associated negatively with infant mortality but do not have expected correlation with income. The links to institutions from cultural variables, although statistically significant, are of minor importance. As expected, the level of urbanization is positively linked to income, agricultural output and to coastal location. Finally, per capita income is strongly negatively affected by infant mortality and positively impacted by urbanization.

The results for the spatial system of equations documented in Table 4 shows results that are broadly similar to those in Table 3, but allowing for spatial dependence changes the results in important ways. First, after controlling for the observed variables we find significant spatial lags among all of the endogenous variables, except for infant mortality. Only infant mortality is explained by our data on country characteristics, without recourse to unobserved neighborhood effects. For agricultural output, schooling, institutional quality, urbanization, and international trade there are positive spatial lags, while for income there is a small negative spatial lag. That is, when controlling for the positive neighborhood effects in these endogenous determinants of income, the remaining influences are negatively correlated across space. Since our model is linear, this result could be due to diminishing returns to these or other inputs.

Having controlled for unobserved spillovers and regional characteristics, the measured variables shown in Table 4 show several very interesting correlations. First, for agricultural output, our variables on land quality and quantity, prevalence of winter frosts and malaria ecology remain significant and of the expected sign. Total rainfall is not significant, and there is a positive time trend as shown by increasing coefficients on the period dummies. In the second column, for infant mortality, both malaria ecology and income are significant as expected. The residual effect of time is quite large and significant, suggesting important technological improvements allowing lower infant mortality at a given level of income and malaria ecology. The third column, for schooling, shows both infant mortality and income to be significant and of the expected sign, with no residual time trend. The fourth variable, institutional quality, has a positive correlation with income, and a small correlation with social history as measured by percent Catholic and Muslim, with no residual time trend. Urbanization is correlated with local agricultural output (though weakly), with income and with coastal location, and has a small positive time trend when controlling for these factors. Trade is negatively correlated with population size and with income (perhaps due to the increased role of non-traded services), and positively linked to coastal location, with a small positive time trend. And in our final equation, all of these endogenous variables have independent correlation with income, except for agriculture output and schooling. In other words, exogenously higher agricultural output drives increased income by facilitating urbanization, and in this specification increased schooling is a result but not a cause of income growth. There is also a large residual effect of time on real income, with unmeasured factors driving increases in measured income from period to period from 1960 until 1975, followed by decreases through 2000.

In sum, when controlling for spatial processes in this model, we maintain support for both the "institutionalist" and "geographic" schools of thought. Geographic factors such as malaria ecology, coastal location and seasonal frost are found to have significant independent effects on the system, influencing institutional quality but not completely determining it, and a country's institutional quality then does have a strong independent role in income.

5. Conclusions

This paper uses panel data in a system of simultaneous equations, controlling for spatial spillovers and unobserved spatial heterogeneity, to explore how measured country characteristics such as physical geography and institutions might be linked to real income per person. This approach offers a new kind of test for how particular types of technologies and institutions might affect income, and then test the robustness of each variable against various kinds of neighborhood effects.

The endogenous variables associated with income are agricultural output per capita (as measured by the FAO), health status (as measured by infant mortality), educational attainment (as measured by years of schooling), institutional quality (as measured by a combination of Freedom House and ICRG indexes), and urbanization (percentage of the population in towns or cities). The exogenous variables represent climate (which plausibly affect only agricultural output), malaria ecology (only through agriculture or health), social history (only through institutions), coastal location (only through urbanization or trade), and population size (through international trade). With this specification, after controlling for spatial proximity, all of the variables have some independent effect on income, except schooling. This result provides strong empirical support for both "geographic" and "institutionalist" hypotheses. Geographic variables such as land quality, coastal location and malaria prevalence have strong independent effects on income, primarily by facilitating urbanization and declines in infant mortality. Institutional quality also has a strong independent link to income, even when controlling for reverse causality and neighborhood effects.

Most notably, accounting for these country characteristics still leaves large residual spatial lags. This result suggests that our specification has only begun to capture the relevant spillovers and spatial heterogeneity among countries. Understanding these spatial correlations will require more precise measurement of both the unobserved factors driving local agricultural productivity, public health and ease of urbanization, but also more complete accounting for cross-border flows associated with migration, investment or technology diffusion.

Throughout the paper we have indicated potential extensions and variations to be addressed in future work. Among those are testing for exogeneity and exclusion restrictions, the incorporation of heteroskedasticity following the procedures developed in Kelejian and Prucha (2005), an assessment of parameter heterogeneity and other robustness checks, and consideration of the temporal dynamics of the system.

References

Abreu, M., H.L.F. de Groot and R.J.G.M. Florax (2005a) Space and Growth: A Survey of Empirical Evidence and Methods, *Région et Développement* **21**: 13–44.

Abreu, M., H.L.F. de Groot and R.J.G.M. Florax (2005b) A Meta-Analysis of β-Convergence: The Legendary 2%, *Journal of Economic Surveys* **19**(3): 389–420.

Acemoglu, D., S. Johnson and J. Robinson (2001) The Colonial Origins of Comparative Development, *American Economic Review* **91**(5): 1369–1401.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer.

Anselin, L. (1996) The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association, in: M. Fischer, H. Scholten and D. Unwin (eds.), *Spatial Analytical Perspectives on GIS*, London: Taylor and Francis.

Anselin, L. (2003) Spatial Externalities, Spatial Multipliers, and Spatial Econometrics, *International Regional Science Review* **26**(2): 153–66.

Anselin, L. and H. Kelejian (1997) Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors, *International Regional Science Review* **20**(1&2): 153–82.
Anselin, L., A. Bera, R. Florax and M. Yoon (1996) Simple Diagnostic Tests for Spatial Dependence, *Regional Science and Urban Economics* **26**(1): 77–104.
Anselin, L., J. Le Gallo and H. Jayet (2006, forthcoming) Spatial Panel Econometrics, in L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, Dordrecht: Kluwer.
Barro, R.J. (1991) Economic Growth in a Cross Section of Countries, *The Quarterly Journal of Economics* **106**(2): 407–43.
Barro, R.J. and J.-W. Lee (2001) International Data On Educational Attainment: Updates And Implications, *Oxford Economic Papers* **53**(3): 541–63.
Baumol, W.J. (1986) Productivity Growth, Convergence, and Welfare: What the Long-run Data Show, *American Economic Review* **76**(5): 1072–85.
Bell, K. and N. Bockstael (2000) Applying the Generalized Moments Estimation Approach to Spatial Problems Involving Microlevel Data, *The Review of Economics and Statistics* **82**: 72–82.
Caselli, F., G. Esquivel and F. Lefort (1996) Reopening the Convergence Debate: A New Look at Cross-country Growth Empirics, *Journal of Economic Growth* **1**(3): 363–89.
Cho, D. (1996) An Alternative Interpretation of Conditional Convergence Results, *Journal of Money, Credit, and Banking* **28**(4): 669–81.
Cliff, A. and J. Ord (1981) *Spatial Processes, Models, and Applications*, London: Pion.
Conley, T.G. and E. Ligon (2002) Economic Distance and Cross-Country Spillovers, *Journal of Economic Growth* **7**(2): 157–87.
Diamond, J.M. (1997) *Guns, Germs and Steel: The Fate of Human Societies*, New York: Norton.
Dobson, S., C. Ramlogan and E. Strobl (2003) Why Do Rates of β-Convergence Differ? A Meta-Regression Analysis, *Scottish Journal of Political Economy* **53**(2): 153–73.
Easterly, W. and R. Levine (2003) Tropics, Germs, and Crops: How Endowments Influence Economic Development, *Journal of Monetary Economics* **50**(1): 3–39.
Ertur, C. and W. Koch (2005) Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, paper presented at the 45th conference of the European Regional Science Association, Amsterdam.
FAO, Food and Agriculture Organization (2004) FAOStat, http://faostat.fao.org and unpublished output series from Jan Poulisse (pers. comm.), Rome: Food and Agriculture Organization of the United Nations.
Fingleton B. and E. López-Bazo (2006) Empirical Growth Models with Spatial Effects, *Papers in Regional Science* **85**(2): 177–98.
Freedom House (2005) Freedom in the World Country Ratings, http://www.freedomhouse.org.
Godfrey, L.G. (1988) *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*, Cambridge: Cambridge University Press.
Gundlach, E. (2004) The Primacy of Institutions Reconsidered: The Effects of Malaria Prevalence in the Empirics of Development, Kiel Institute for World Economics, Working Papers 1210.
Hall, R. and C.I. Jones (1999) Why Do Some Countries Produce So Much More Output per Worker than Others?, *Quarterly Journal of Economics* **114**(1): 83–116.
ICRG (2006) Country data, *International Country Risk Guide*, http://www.countrydata.com

IPCC, International Panel on Climate Change (1999) IPCC Data Distribution Data Centre CDROM, April 1999, http://ipcc-ddc.cru.uea.ac.uk.
Kelejian, H. and D. Robinson (1993) A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model, *Papers in Regional Science* **72**: 297–312.
Kelejian, H.H. and I.R. Prucha (1999) A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review* **40**(2): 509–33.
Kelejian, H.H. and I.R. Prucha (2001) On the Asymptotic Distribution of the Moran I test Statistic with Applications, *Journal of Econometrics* **104**(2001): 219–57.
Kelejian, H.H. and I.R. Prucha (2004) Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations, *Journal of Econometrics* **118**(1-2): 27–50.
Kelejian, H.H. and I.R. Prucha (2005) Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances, manuscript, University of Maryland.
Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S.E. Sachs, and J. Sachs (2004), A Global Index Representing the Stability of Malaria Transmission, *American Journal of Tropical Medecine and Hygiene* **70**: 486-498.
Klein, R. and F. Vella (2005) Estimating a Class of Triangular Simultaneous Equations Models without Exclusion Restrictions, Centre for Microdata Methods and Practice, Working Papers, CWP08/05.
Lee, L.-F. (2003) Best Spatial Two-Stage Least Squares Estimation for a Spatial Autoregressive Model with Autoregressive Disturbances, *Econometric Reviews* **22**: 307–35.
Magrini, S. (2004) Regional (Di)Convergence, in: V. Henderson and J.-F. Thisse (eds.), *Handbook of Urban and Regional Economics*, Amsterdam: Elsevier.
Masters, W.A. and M.S. McMillan (2001) Climate and Scale in Economic Growth, *Journal of Economic Growth* **6**(3): 167–86.
Masters, W.A. and K. Wiebe (2000) Climate and Agricultural Productivity, paper presented at the Conference on Biophysical Constraints in Tropical Agriculture, Harvard University.
Matsuyama, K. (1992) Agricultural Productivity, Comparative Advantage, and Economic. Growth. *Journal of Economic Theory* **58**(2): 317–334.
Mellor, J.W. and B.F. Johnston (1961), The Role of Agriculture in Economic Development. *American Economic Review* **51** (4, Sep.): 566-593.
Mitchell**,** T.D., T.R. Carter, P.D. Jones, M. Hulme and M. New (2003) A Comprehensive Set of High-Resolution Grids of Monthly Climate for Europe and the Globe: The Observed Record (1901– 2000) and 16 Scenarios (2001– 2100), *Journal of Climate*, http:// earthscape.org.
North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*, New York: Cambridge University Press.
NRCS, Natural Resources Conservation Service (1999) World Soil Resources database http://www.nhq.nrcs.usda.gov/WSR/, Washington, DC: Natural Resources Conservation Service, USDA.
Rodrik, D., A. Subramanian and F. Trebbi (2004) Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development, *Journal of Economic Growth* **9**(2): 131–65.
Sachs, J.D. (2001) Tropical Underdevelopment, NBER Working Paper 8119, http://www.nber. org/papers/W8119.

Sala-i-Martin, X. (1997) I Just Ran Two Million Regressions, *American Economic Review* **87**(2): 178–83.
Sala-i-Martin, X. (2006) The World Distribution of Income, *Quarterly Journal of Economics* **121**(2): 351–97.
Temple, J. (1999) The New Growth Evidence, *Journal of Economic Literature* **37**: 112–56.
Tiefelsdorf, M. (2000) *Modelling Spatial Processes: The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I*, Heidelberg: Springer-Verlag.
USGS, United States Geological Survey (1999) Global Land Cover Characterization, U.S. Geological Survey, the University of Nebraska-Lincoln, and the Joint Research Centre of the European Commission, http://edcwww.cr.usgs.gov/ landdaac/glcc/ glcc.html.

Table 1. Descriptive statistics[a]

Variable[b]\statistic	Mean	Variance	Minimum	Maximum	Skewness	Kurtosis
Agricultural output	0.254	0.060	0.001	1.744	2.724	12.257
Agricultural land	0.412	0.204	0.001	3.731	3.751	20.687
Land quality	20.340	370.685	0.001	76.296	1.029	3.292
Frost	0.366	0.194	0.000	1.000	0.533	1.423
Rainfall	1278.946	575948.400	46.233	3416.267	0.616	2.561
Malaria	3.713	40.247	0.000	30.095	2.026	6.625
Population (× 1000)	37.828	16100.000	0.042	1249.134	6.635	50.647
Infant mortality	72.047	3052.841	2.980	285.000	0.671	2.647
Trade	62.756	2651.960	5.048	541.396	3.388	23.463
Income (× 1000)	6.407	42.500	0.384	33.711	1.469	4.476
Schooling	4.684	8.014	0.120	12.050	0.423	2.407
Institutional quality	0.403	0.086	0.143	1.000	1.143	2.856
Catholic	38.063	1406.900	0.000	96.900	0.489	1.582
Protestant	14.122	478.400	0.000	97.800	2.138	7.397
Muslim	17.886	968.232	0.000	99.700	1.769	4.550
Urbanization	45.183	601.941	2.230	100.000	0.238	2.154
Coastal	48.138	1451.542	0.001	100.000	0.189	1.464
Time dummies[c]	0.000	0.222	–1.000	1.000	0.000	4.500

[a] Based on 95 countries, five-year averages from 1960 through 2000.

[b] Variable definitions are detailed in the text and summarized here: *agricultural output* is an index of net farm production per capita at international prices in 2000 US dollars; *agricultural land* is land used in agriculture in thousands of hectares; *land quality* is the percentage of a country's farmland that falls in the top three categories of fertility; *frost* refers to the proportion of a country's land receiving five or more frost days per month in winter; *rainfall* is average total annual precipitation over the country's land mass in millimeter; *malaria* is an ecological index of malaria transmissibility; *infant mortality* is per 1,000 live births; *income* is real GDP per capita in PPP terms, expressed in 2000 US dollars; *schooling* is the average number of years of education for the population over age 15; *institutional quality* is the average of IRCG indexes for "corruption", "military in politics", "religion in politics", "law and order", "democratic accountability", and "bureaucratic quality", combined with Freedom House indexes for "Political Rights" and "Civil Liberty"; *catholic*, *protestant* and *muslim* are estimated percentages of the population with the specified religion; *urbanization* is the percentage of the population living in urban areas; *coastal* is the percentage of a country's land that is within 100 km of a seacoast or navigable river.

[c] The time dummies allow fixed effects for 1960, 1965, etc. and are subsequently recomputed as deviations from the omitted category, 1960.

Table 2. Regression output, equation-by-equation estimation, OLS with diagnostics for spatial effects[a,b]

Variables	Agricultural output	Infant mortality	Schooling	Institutional quality	Urbanization	Trade	Income
Agricultural output					–0.041*		0.103***
					(0.023)		(0.022)
Agricultural land	0.397***						
	(0.025)						
Land quality	0.052***						
	(0.010)						
Frost	0.105***						
	(0.031)						
Rainfall	0.072***						
	(0.015)						
Malaria	–0.092***	0.061***					
	(0.011)	(0.006)					
Infant mortality			–0.436***	–0.247***			–0.606***
			(0.039)	(0.033)			(0.026)
Income		–0.689***	0.201***	0.224***	0.458***	–0.115***	
		(0.019)	(0.036)	(0.030)	(0.019)	(0.024)	
Schooling							–0.099***
							(0.028)
Institutional quality							0.230***
							(0.033)
Trade							–0.057***
							(0.019)
Catholic				0.001*			
				(0.0004			
Protestant				0.005***			
				(0.001)			
Muslim				–0.001			
				(0.001)			
Urbanization							0.516***
							(0.027)
Coastal					0.056***	0.043***	
					(0.005)	(0.007)	
Population						–0.183***	
						(0.013)	
D_{1965}	–0.086*	0.300***	–0.092*	0.113***	–0.121***	–0.286***	0.114***
	(0.053)	(0.042)	(0.051)	(0.039)	(0.042)	(0.063)	(0.038)
D_{1970}	–0.045	0.249***	–0.054	0.044	–0.085**	–0.161***	0.131***
	(0.053)	(0.041)	(0.051)	(0.039)	(0.042)	(0.062)	(0.037)
D_{1975}	–0.014	0.208***	–0.009	–0.018	–0.049	–0.079	0.146***
	(0.053)	(0.041)	(0.051)	(0.039)	(0.041)	(0.062)	(0.037)
D_{1980}	0.004	0.034	0.017	–0.048	–0.003	0.022	0.056
	(0.053)	(0.041)	(0.050)	(0.038)	(0.041)	(0.062)	(0.037)
D_{1985}	0.022	–0.095**	0.048	–0.057	0.051	0.009	–0.034
	(0.053)	(0.041)	(0.050)	(0.038)	(0.041)	(0.062)	(0.037)
D_{1990}	0.042	–0.224***	0.075	–0.041	0.091**	0.138**	–0.112***
	(0.053)	(0.041)	(0.051)	(0.039)	(0.042)	(0.062)	(0.037)
D_{1995}	0.079	–0.345***	0.068	–0.080**	0.126***	0.295***	–0.173***
	(0.053)	(0.041)	(0.052)	(0.040)	(0.042)	(0.062)	(0.038)
D_{2000}	0.120**	–0.442***	0.041	–0.069*	0.148***	0.397***	–0.227***
	(0.053)	(0.042)	(0.053)	(0.041)	(0.042)	(0.063)	(0.039)
Constant	–1.884***	9.625***	1.305***	–2.117***	–0.371**	6.422***	9.488***
	(0.187)	(0.152)	(0.440)	(0.351)	(0.178)	(0.233)	(0.161)

Table 2. Continued

Variables	Agricultural output	Infant mortality	Schooling	Institutional quality	Urbanization	Trade	Income
Condition number	32	32	66	70	35	29	36
Jarque-Bera	185.89***	12.39***	351.27***	0.89	362.63***	39.50***	26.62***
Breusch-Pagan	49.43***	30.29***	130.99***	26.44	85.71***	75.99***	77.88***
Moran's I	2.65***	–3.18***	–6.04***	–0.62	6.99	3.13***	–0.004
LM-error	2.94*	13.97***	41.09***	1.75	33.54***	4.69**	0.58
Robust LM-error	0.0004	6.28**	10.53***	13.70***	0.78	0.63	0.98
LM-lag	4.40**	10.01***	34.59***	5.94**	38.89***	4.08**	15.57***
Robust LM-lag	1.47	2.32	4.03***	17.90***	6.13**	0.02	15.97***
SARMA	4.40	16.29***	45.12***	19.64***	39.67***	4.71*	16.55***
R^2-adjusted	0.47	0.83	0.63	0.63	0.67	0.24	0.87
AIC	1410.36	980.90	1306.21	849.66	988.50	1678.76	790.47
Log-likelihood	–691.18	–479.45	–642.10	–410.83	–482.25	–827.38	–380.23

[a] All variables enter in logarithmic form, except for the time dummies and the constant. Significance is indicated by ***, ** and * for the 1, 5, and 10 per cent level, respectively, with standard errors in parentheses.

[b] The Jarque-Bera and the Breusch-Pagan tests are asymptotically χ^2 distributed, and test for normality of the errors and homoskedasticity with random coefficient variation as the alternative hypothesis, respectively. In cases where the null hypothesis of the Jarque-Bera test is rejected, the Koenker-Basett variant instead of the Breusch-Pagan version is reported. For details on the spatial misspecification tests see Anselin et al. (1996).

Table 3. Regression output, system estimation, 3SLS not allowing for spatial spillovers[a,b]

Variables	Agricultural output	Infant mortality	Schooling	Institutional quality	Urbanization	Trade	Income
Agricultural output					0.063**		0.077***
					(0.032)		(0.028)
Agricultural land	0.434***						
	(0.023)						
Land quality	0.058***						
	(0.008)						
Frost	0.043***						
	(0.009)						
Rainfall	0.023						
	(0.031)						
Malaria	–0.076***	0.036***					
	(0.008)	(0.007)					
Infant mortality			–0.979***	–0.403***			–0.860***
			(0.075)	(0.092)			(0.068)
Income		–0.781***	–0.208***	0.102	0.409***	–0.244***	
		(0.028)	(0.071)	(0.082)	(0.029)	(0.033)	
Schooling							–0.386***
							(0.053)
Institutional quality							0.320***
							(0.093)
Trade							–0.061**
							(0.028)
Catholic				0.001			
				(0.001)			
Protestant				0.004***			
				(0.001)			
Muslim				–0.001**			
				(0.001)			
Urbanization							0.458***
							(0.047)
Coastal					0.060***	0.070***	
					(0.005)	(0.008)	
Population						–0.197***	
						(0.013)	
D_{1965}	–0.089	0.274***	0.062	0.155***	–0.136***	–0.326***	0.127***
	(0.057)	(0.042)	(0.054)	(0.045)	(0.044)	(0.062)	(0.047)
D_{1970}	–0.045	0.236***	0.077	0.081*	–0.094**	–0.183***	0.149***
	(0.057)	(0.041)	(0.053)	(0.044)	(0.044)	(0.061)	(0.043)
D_{1975}	–0.01	0.206***	0.104**	0.014	–0.052	–0.083	0.175***
	(0.057)	(0.041)	(0.053)	(0.043)	(0.043)	(0.061)	(0.042)
D_{1980}	0.000	0.04	0.038	–0.042	0.000	0.031	0.068***
	(0.057)	(0.041)	(0.051)	(0.039)	(0.043)	(0.061)	(0.041)
D_{1985}	0.017	–0.085**	0.001	–0.071*	0.058	0.025	–0.031
	(0.057)	(0.041)	(0.051)	(0.040)	(0.043)	(0.061)	(0.042)
D_{1990}	0.042	–0.209***	–0.041	–0.074*	0.102**	0.162***	–0.122***
	(0.057)	(0.041)	(0.053)	(0.043)	(0.044)	(0.061)	(0.043)
D_{1995}	0.083	–0.324***	–0.112	–0.130***	0.138***	0.328***	–0.199***
	(0.057)	(0.042)	(0.056)	(0.048)	(0.044)	(0.062)	(0.046)
D_{2000}	0.130**	–0.416***	–0.189	–0.134**	0.161***	0.440***	–0.277***
	(0.057)	(0.042)	(0.058)	(0.053)	(0.044)	(0.062)	(0.048)
Constant	–1.372***	10.347***	6.769	–0.475	0.195	7.552***	11.128***
	(0.190)	(0.227)	(0.867)	(0.995)	(0.276)	(0.290)	(0.332)

Table 3. Continued

Variables	Agricultural output	Infant mortality	Schooling	Institutional quality	Urbanization	Trade	Income
R^2	0.50	0.83	0.55	0.62	0.67	0.25	0.84
Goodness-of-fit	942.64***	3534.99***	1407.15***	1178.58***	1389.49***	360.02***	3554.23***

[a] See footnote a to Table 2.

[b] Note that the R^2-value is not restricted to the usual [–1,+1] interval. The goodness-of-fit test is a Wald test with an asymptotic χ^2-distribution.

Table 4. Regression output, system estimation, full information estimator for the ARAR specification[a]

Variables	Agricultural output	Infant mortality	Schooling	Institutional quality	Urbanization	Trade	Income
$W_{\text{Agricultural output}}$	0.501***						
	(0.042)						
$W_{\text{Infant mortality}}$		0.018					
		(0.036)					
$W_{\text{Schooling}}$			0.571***				
			(0.053)				
$W_{\text{Institutional quality}}$				0.273***			
				(0.054)			
$W_{\text{Urbanization}}$					0.422***		
					(0.040)		
W_{Trade}						0.698***	
						(0.096)	
W_{Income}							–0.077**
							(0.030)
Agricultural output					0.046*		0.031
					(0.026)		(0.020)
Agricultural land	0.368***						
	(0.021)						
Land quality	0.044***						
	(0.008)						
Frost	0.032**						
	(0.009)						
Rainfall	0.028						
	(0.028)						
Malaria	–0.044***						
	(0.009)						
Infant mortality			–0.180**	–0.033			–0.617***
			(0.052)	(0.055)			(0.040)
Income		–0.814***	0.221***	0.346***	0.310***	–0.064**	
		(0.028)	(0.053)	(0.046)	(0.027)	(0.024)	
Schooling		0.032***					0.037
		(0.005)					(0.037)
Institutional quality							0.438***
							(0.052)
Trade							–0.049*
							(0.023)
Catholic				0.004***			
				(0.001)			
Protestant				–0.001			
				(0.001)			
Muslim				–0.002**			
				(0.001)			
Urbanization							0.434***
							(0.035)
Coastal					0.039***	0.036***	
					(0.004)	(0.007)	
Population						–0.158***	
						(0.013)	

Table 4. Continued

Variables	Agricultural output	Infant mortality	Schooling	Institutional quality	Urbanization	Trade	Income
D_{1965}	–0.110*	0.179***	–0.001	0.057	–0.055	–0.169**	0.092*
	(0.055)	(0.040)	(0.051)	(0.040)	(0.041)	(0.067)	(0.038)
D_{1970}	–0.061	0.156***	–0.005	0.006	–0.039	–0.093	0.112**
	(0.055)	(0.040)	(0.050)	(0.039)	(0.040)	(0.063)	(0.037)
D_{1975}	–0.023	0.140**	–0.003	–0.042	–0.025	–0.042	0.131***
	(0.055)	(0.039)	(0.049)	(0.038)	(0.040)	(0.061)	(0.037)
D_{1980}	0.003	0.028	–0.004	–0.046	–0.003	0.013	0.053
	(0.055)	(0.039)	(0.048)	(0.037)	(0.040)	(0.060)	(0.037)
D_{1985}	0.031	–0.056	0.005	–0.036	0.025	0.022	–0.025
	(0.055)	(0.039)	(0.048)	(0.037)	(0.040)	(0.060)	(0.037)
D_{1990}	0.063	–0.137**	0.013	–0.009	0.044	0.093	–0.097**
	(0.055)	(0.040)	(0.050)	(0.039)	(0.040)	(0.062)	(0.037)
D_{1995}	0.105*	–0.214***	0.003	–0.019	0.059	0.168*	–0.146**
	(0.055)	(0.041)	(0.052)	(0.041)	(0.041)	(0.068)	(0.038)
D_{2000}	0.144**	–0.274***	–0.018	–0.005	0.065*	0.217**	–0.192***
	(0.055)	(0.041)	(0.053)	(0.042)	(0.041)	(0.075)	(0.038)
Constant	–0.592**	10.555***	–0.566	–3.529***	–0.522*	3.043***	10.477***
	(0.179)	(0.325)	(0.601)	(0.558)	(0.225)	(0.509)	(0.314)
Implicit λ[b]	–0.315**	0.304***	0.004	0.044	0.020	–0.324	0.214
	(0.111)	(0.016)	(0.036)	(0.093)	(0.048)	(0.286)	(0.150)

[a] See footnote a to Table 2.
[b] Estimated values and standard errors for λ based on the GM estimator in the second step of the estimation procedure; used in the Cochrane-Orcutt transformation to obtain full information estimates.

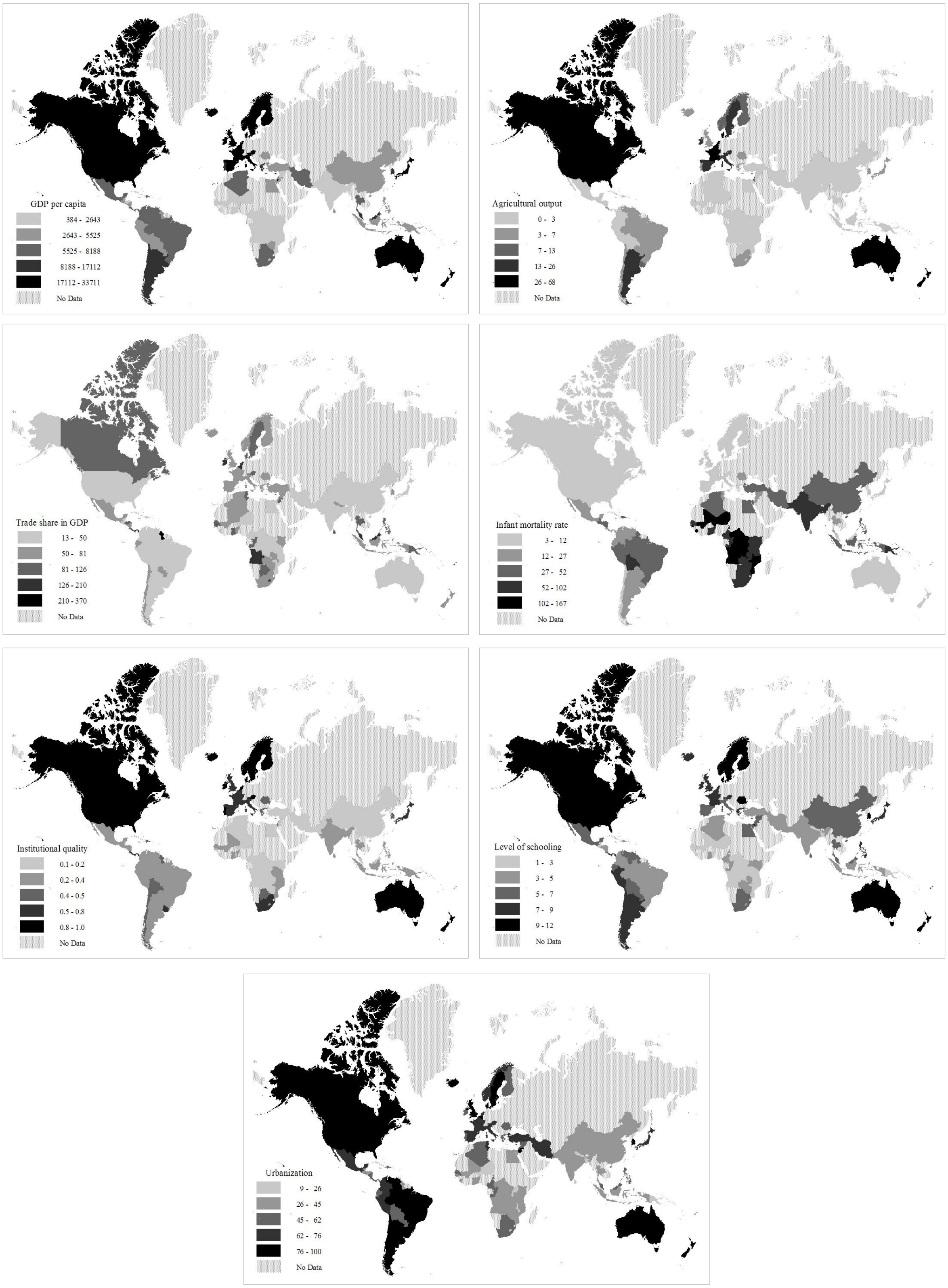

Figure 1. Maps of GDP per capita, agricultural output, trade share, infant mortality rate, institutional quality, level of schooling, and urbanization, in 2000

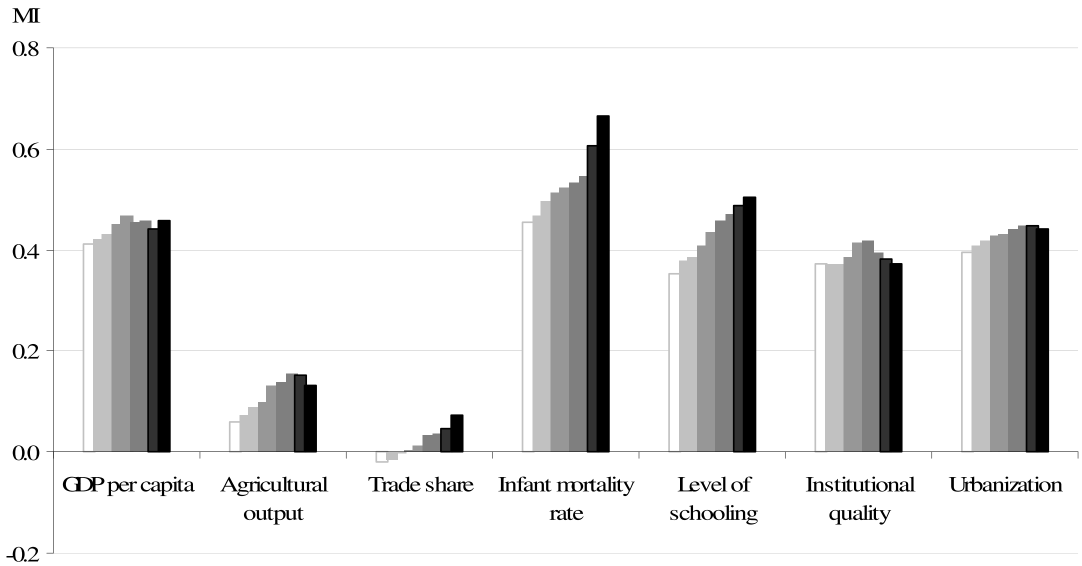

Figure 2. Moran's *I* grouped by dependent variable from 1960 (left) through 2000 (right)

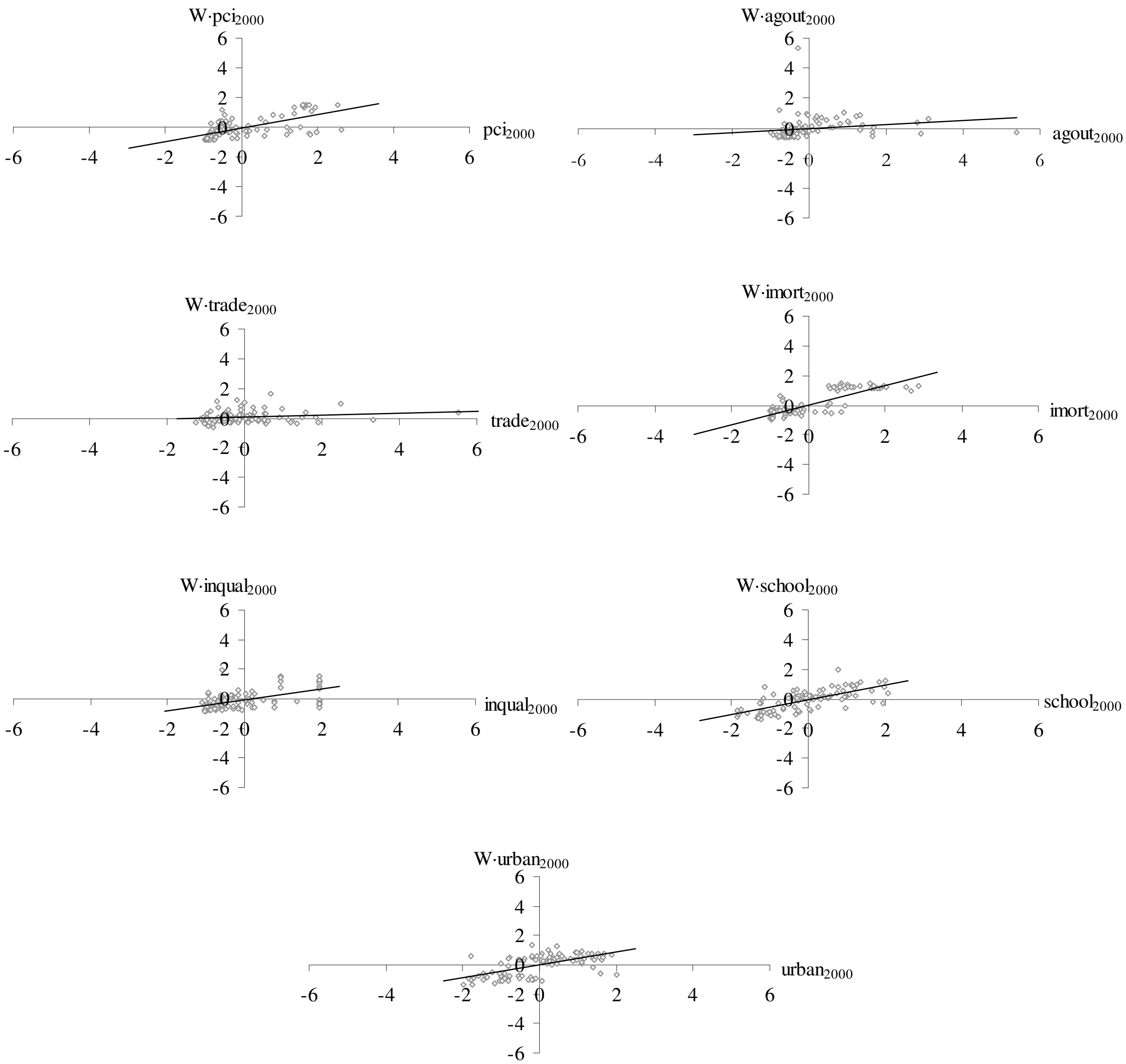

Figure 3. Moran scatterplots for GDP per capita, agricultural output, trade share, infant mortality rate, institutional quality, level of schooling, and urbanization, in 2000

Appendix

Table A1. Countries included in the sample

Algeria	Ghana	Pakistan
Angola	Greece	Panama
Argentina	Guatemala	Papua New Guinea
Australia	Guinea-Bissau	Paraguay
Austria	Guyana	Peru
Bangladesh	Haiti	Philippines
Barbados	Honduras	Portugal
Belgium	Iceland	Romania
Benin	India	Rwanda
Bolivia	Indonesia	Senegal
Botswana	Iran	Seychelles
Brazil	Ireland	Sierra Leone
Burundi	Israel	Singapore
Cameroon	Italy	South Africa
Canada	Jamaica	Spain
Central African Republic	Japan	Sri Lanka
Chile	Jordan	Sweden
China	Kenya	Syria
Colombia	Korea, Republic	Tanzania
Congo, Democratic Republic	Lesotho	Thailand
Congo, Republic of Congo	Malawi	Togo
Costa Rica	Malaysia	Trinidad and Tobago
Cyprus	Mali	Tunisia
Denmark	Mauritius	Turkey
Dominican Republic	Mexico	Uganda
Ecuador	Mozambique	United Kingdom
Egypt	Nepal	United States
El Salvador	Netherlands	Uruguay
Fiji	New Zealand	Venezuela
Finland	Nicaragua	Zambia
France	Niger	Zimbabwe
Gambia, The	Norway	